1-10-2017

Validating Imaginary Worlds? The AdViSHE Assessment Tool

Paul C. Langley

University of Minnesota, langley@maimonresearch.com

Validating Imaginary Worlds? The AdViSHE Assessment Tool

Paul C Langley, PhD
Adjunct Professor, College of Pharmacy, University of Minnesota

Abstract
The publication in April 2016 of the Assessment of the Validation Status of Health-Economic Decision Models (AdViSHE) checklist for decision models raises a number of issues that the health technology assessment literature has yet to address. The principal issue being the role of decision models in generating claims that are evaluable and replicable. Unfortunately, this is not addressed in this new checklist which is intended to address the perceived need for a tradeoff between confidence in a decision model and the need to allocate resources by developers and payers to validating the model. Irrespective of the degree of confidence a developer or payers may have in the sufficiency of the model in representing 'reality' unless the model has generated evaluable claims and evidence for those claims in target treating populations, the model fails the standards of normal science. Apart from the absence of a commitment in the AdViSHE checklist to the modeling of claims that are evaluable and replicable, the validation check list makes no allowance for a product pricing strategy that may commits a manufacturer to regular and substantial annual or semi-annual product price increases. Indeed, product pricing assumptions are conspicuous by their absence. The commentary argues that failure to accommodate anticipated pricing behavior renders lifetime cost-per-QALY models and the application of willingness-to-pay thresholds meaningless.

Introduction

In April 2016, Vemer et al published details of the Assessment of the Validation Status of Health-Economic Decision Models (AdViSHE) tool [1]. This tool is a checklist to support the validation of decision models. The objectives are to evaluate 'whether a model is a proper and sufficient representation of the system it is intended to represent'. For the authors, a 'proper' model is one that 'is in accordance with what is known about the system' while 'sufficient' means that the results of the model can serve as a 'solid basis' for decision making'. Unlike previous attempts to set standards for model validity and quality assessment, the AdViSHE tool focuses on the potential trade-off between building confidence in the model and scare resources to support its validation. The tool presents a prioritized list of validation efforts with the objective of saving on scare resources yet improving the model's validation status and acceptability for developers and payers.

The purpose of this commentary is to consider the AdViSHE tool kit and its recommendations for setting validation priorities from the perspective that the validation assessment is only meaningful if the model supports claims for product outcomes that are credible, evaluable and replicable, and where there is evidence for claims assessment [2]. If the tool kit fails to support the assessment of claims, irrespective of what may be seen as the intrinsic merits or 'sufficiency' of the model, then the model fails the standards of normal science [3][4][5][6]. As such, it should be considered as 'pseudoscience', sharing the stage with intelligent design rather than natural selection. If this is the case then regardless of the belief in the 'sufficiency of the 'proper' model, where the results are necessarily entailed, it is difficult to see its role in providing a meaningful input to inform formulary decisions. Input that includes providing feedback for the claims made and supporting ongoing disease area and therapeutic class reviews. Irrespective, therefore, of the confidence the model builder may have in the model, this may not be shared by payers [7][8]. More to the point, payers should be advised that models generating non-evaluable product should be rejected.

The AdViSHE Toolkit

The toolkif is the outcome of what is described as an exhaustive review process of modified Delphi rounds and a workshop at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 2014 Montreal meeting. The agreed final version of AdViSHE consists of 13 questions covering the validation of: (i) the conceptual model; (ii) input

Corresponding author: Paul C Langley, PhD
Adjunct Professor, College of Pharmacy
University of Minnesota
Director, Maimon Research LLC
5061 North Apache Hills Trail, Tucson, AZ 85750
Tel: 520-577-0436; Email: langley@maimonresearch.com
Web:www.maimonresearch.net

data; (iii) the computerized model; and (iv) the behavior and accuracy of the model outcomes.

The assessment checklist is in five parts. Part A addresses the issue of the validation of the conceptual model and comprises two questions. The first questions address the question of face validity through asking experts to judge the models appropriateness while the second question asks whether the model has been compared to other conceptual models to establish cross validity. Part B comprises two questions on input data validation to establish, first, an expert assessment of face validity for the appropriateness of the input data and, second, an evaluation of the fit of the model where the input parameters are based upon regression models. Part C comprises four questions on the validation of the computerized model. These cover: (i) an expert external review of the model; (ii) model testing for extreme values; (iii) a logic assessment of patient tracking through the model; and (iv) testing of sub-modules in the model. Part D considers, with four questions, the operational validity of the model. These are: (i) expert assessment of the face validity of the model outcomes; (ii) cross validation of the outcomes against those that address similar outcomes; (iii) validation against outcomes using alternative input data; and (iv) validation through comparing outcomes to empirical data. A final section, Part E, asks whether other validation techniques have been performed.

The assessment checklist is seen as representing a compromise between what is feasible and what is necessary in decision modeling from developer and payer perspectives. It is seen as supplementing existing validation tools, with particular reference to the ISPOR-SMDM modeling standards, in asking which validation aspects were tested, how they were tested and where outcomes are reported [9] [10]. In seeking to avoid duplicating validations while identifying unreported validation standards, the benefit of the AdViSHE tool is seen by its authors in its allowing model developers to build confidence in their model through commenting on validations already undertaken. As such, it reduces overlap between validation of model developers and those of model users. The key assumption being, particularly from a payer perspective, that the criteria are relevant to their decision making.

Credibility, Evaluation and Replication
Although the AdViSHE framework addresses the issue of outcomes, there is no attempt to raise the issue of evaluable and replicable claims as a criterion for model validation. This is major oversight because it allows the model builder to fall back on claims that if the model is considered sufficient in its representation of 'reality' then, because the simulated outcomes are necessarily entailed, the model can 'inform'

decision makers. Unfortunately, this position perpetuates the acceptance in health technology assessment of models that claim to meet validation standards yet fail to meet the standards for credible, evaluable and replicable claims. This is seen in the ISPOR-SDMS standards for modeling where predictive validation is seen as perhaps the most important test for model credibility as a validation standard yet one that is considered neither necessary nor sufficient in judging the merits of a model.

It is far from clear what the term 'outcomes' is supposed to encompass in the AdViSHE checklist. While experts (Question D1) are asked to judge the appropriateness of the model outcomes, there is no discrimination between evaluable and non-evaluable claims. The same criticism applies to cross validation (Question D2) where the model outcomes are to be assessed against those of other models that address similar outcomes. There is no requirement that claims from competing models should be evaluated empirically. If there is no attempt to present evaluable claims, then the validation assessment is asked to contrast one set of non-evaluable claims against another. This seems to be an odd form of validation where a more appropriate assessment, if there are evaluable claims, is to contrast one model's claims against those of another empirically. If this is not done than, irrespective of claims for the superiority of one model over another in its structure, assumptions, etc. the health care decision maker has no idea whether the claims are right or even if they are wrong. The same argument applies in respect of Question D3 where outcomes are validated against alternative input data. Finally, in validation against empirical data (Question D4) there is still confusion over what form a comparison should take. Information is requested on two aspects of possible empirical assessments: (i) comparisons based on summary statistics or patient level data sets; and (ii) differences between model outcomes and empirical data. These comparisons apparently involve (i) a comparison against the data sources on which the model is based (dependent validation) and (ii) a comparison against a data source that was not used to build the model (independent validation). While the latter comparison could be interpreted as an evaluation of claims in target treating populations, the context is far from clear. There is no hint that decision makers may prefer models that actually generate evaluable claims.

Lifetime Cost-Per-QALY Models
Although not mentioned, the AdViSHE toolkit is presumably relevant, not only to models that are designed to generate evaluable and replicable claims for outcomes in target populations, but also for models that are best described as 'imaginary worlds'. The failure to make this distinction is important because of the popularity of modeled imaginary

worlds in the health technology assessment literature [11] [12] [13] [14] [15]. It is not clear from the AdViSHE tool whether there is a belief that, within each disease state, there exists an 'ideal' simulation: a model that, on the evidence available, can inform decision makers and justify formulary decisions across competing therapies. Otherwise, in the absence of any evaluable claims, we fall back on a (somewhat pointless) discussion of the relative merits of competing models, jostling for acceptance with competing manufacturers funding and publishing models that support their own product.

Establishing reference case standards for modeled claims in single payer health system does not address the issue of claims credibility. Rather, acceptance of reference case frameworks, notably for long-term or lifetime modeling of chronic disease, reinforces the acceptance of modeled imaginary worlds as a valid input to formulary decision making. In reference case models the focus is on the model itself rather than any assessment of the claims generated by the model. In the case of the UK and the Netherlands, to give two examples, the reference case is the standard [16] [17] [18]. As long as the model meets the reference case criteria and receives, in the case of NICE, the seal of approval from the external review group and the final NICE endorsement, the issue of claims evaluation is irrelevant [19]. Indeed, as pointed out in previous reviews of the NICE evaluation framework, the reference case is not actually intended to generate evaluable claims. Rather, it is a pricing and resource allocation exercise. If, the final version of the model supports cost-per-QALY claims below a lifetime or long term cost-per-QALY willingness to pay threshold then the price proposed by a manufacturer is accepted. If not, negotiations to offer a lower price, discounts or some form of risk contracting ensue. Manufacturers are on notice, therefore, that they need to submit a modeled reference case submission to support approval within the National Health Service. Whether they adjust their target price to meet a willingness-to-pay threshold or opt to argue for a premium 'above threshold' price is their choice. The model, irrespective of how 'sufficient' it is judged is immaterial as it makes no pretensions to generating evaluable claims.

The situation in the US and in other non-single payer health systems is somewhat different. While the view that reference case and similar standards are nothing more than a pricing justification 'rite of passage' is echoed in the US in the reports generated by the Institute for Clinical and Economic Review (ICER), there appears to be little support for cost-per-QALY modeling and willingness-to-pay thresholds [20] [21]. The ICER approach mirrors that of NICE in the application of the reference case cost-per-QALY framework with willingness-to-pay thresholds. Applying threshold values for cost-per-QALY gives a model framework to judge whether or not the price

sought by manufacturers is deemed cost-effective. If the modeled cost-per-QALY falls below a threshold value it is judged cost-effective. Otherwise, the ICER may recommend a discounted price to bring it in line.

A Foot in the Door
Looking back over the past 20 years, it is not difficult to take the view that the effort put into developing standards and validation criteria for modeling and study design is seen, from the perspective of manufacturers who have underwritten much of this activity, as nothing more than an 'academically respectable' support for pricing and market share strategies. Unfortunately, the AdViSHE tool does nothing to dispel this belief as pricing assumptions are not considered as a validation element.

Of course, as noted, in single payer systems, caps can be placed on pricing subsequent to market entry. This is not the case in the US. There is abundant evidence for what many observers see as a long term strategy by manufactures for regular price increases over the patent life of the product, supported by coupon discount policies to maintain market share. Indeed, it is difficult (if not impossible) to find lifetime cost-per-QALY models that factor in long term pricing strategies as part of their long term modeled cost-per-QALY claims. Consider the case of multiple sclerosis drugs where a recent study by Hartnung et al, provides estimates of the trend in annual drug costs for nine of the disease modifying therapies (DMTs) from 1993 to 2014 [22]. Apart from the fact that DMT costs are two to three time bigger in the US than other countries, the principal finding is that DMT costs have accelerated well beyond inflation and substantially above rates for drugs observed in a similar biologic class. Annualized change in the cost of the DMTs ranged from 35.7% (glatiramer acetate) to 7.9% for fingolimod. Four of the DMTs had annualized cost increases above 20% and four with annualized price increases between 13.0% and 16.8%. Natalizumab, for example, although being withdrawn briefly from the market between February 2005 and June 2006, increased in cost from $25,850 in 2004 to $64,233 in 2013 or an annualized increase of 16.2%.

If long-term pricing strategies are put to one side in favor of an assumption that the market entry WAC of a product will be maintained over its patent lifetime, then it is difficult to see what possible justification there is for a lifetime model to support claims for cost-effectiveness. It is certainly neither 'sufficient' nor 'proper'. Attempts to maintain academic 'purity' through advocating models that mimic the natural course of a disease, willingness-to-pay thresholds and the advocacy (at least in the US) for cost-per-QALY modeling seem misplaced. The AdViSHE tool makes no mention of the advisability of incorporation models that build in unit price

increases as part of the model. Indeed, pricing is not mentioned as a discretionary variable that can be adjusted to support claims for comparative cost-effectiveness.

Indeed, it would not be unreasonable to make the case that failure to include potential long-term pricing strategies in lifetime cost-per-QALY or cost per life year saved models, imparts a substantial element of bias in favor of claims for a manufacturer's product. After all, if we consider the case of multiple sclerosis and consider a pricing strategy that increases an initial WAC by 10% per annum over a ten-year time frame the initial WAC will have increased by 135%. This does not include, of course, potential price increases for other direct medical costs. Given this, it seems a little odd to apply a discount (the standard is 3%) to future costs based on the assumption. direct medical costs, to include, drug prices remain unchanged.

Even if long-term models were modified to accommodate strategic pricing scenarios, the fact remains that such models are not intended to generate evaluable predictions. Until model developers accept the premise that health decision makers require claims that can be validated in a meaningful time frame, such models may be intended to inform but are unlikely to be accepted. In these circumstances it is difficult to see what role the AdViSHE tool can accomplish in bringing the two sides together.

Conclusions
The AdViSHE validation tool is probably best seen as a checklist to support pricing justification models. The fact that the focus is primarily on validating the core structure and assumptions

of the model and not, as has been argued here and on a number of previous occasions, on developing testable hypotheses to support evaluation and replication of clinical and cost-effectiveness claims, makes it unlikely that formulary committees and other payers will pay it much attention. While the AdViSHE checklist is intended as a short-cut to assessing the validity of a model that is submitted to support outcome and pricing claims, the absence of criteria to support the credibility, evaluation and replication of claims is a major oversight

From this perspective of claims evaluation, the willingness of manufacturers to underwrite lifetime cost-per-QALY models should be seen as simply an exercise to justify a pricing strategy. Formulary committees are asked to believe that the model justifies a price consistent with a target formulary position and, if possible, a premium price. The fact that a long-term a strategy of regular price increases renders the initial modeling redundant is beside the point. The issue is one of supporting pricing negotiations and formulary acceptance. A position which is apparently accepted, but possibly not recognized, by academic groups and organizations such as the Academy of Managed Care Pharmacy (AMCP) and ISPOR, together with journal editors. Presumably, it could also be argued that, as long as the cost-per-QALY model is accepted for peer review and publication, the manufacturer has little if any interest in the intrinsic merits of the model or whether or not it adheres to the AdViSHE validation status checklist. The bottom line is achieving formulary acceptance at an entry price consistent with a manufacturer's long-term pricing and market share strategy.

References

[1] Vemer PO, Ramos IC, van Voorn GAK et al. AdViSHE: A validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics*. 2016;34:340-61

[2] Langley PC. Supporting formulary decisions: The discovery of new facts or constructed evidence? *Inov Pharm*. 2016; 7(2): No.15

[3] Langley P. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-999

[4] Wootton D. *The Invention of Science*. New York: Harper, 2015

[5] Langley P. The status of modeled claims. *J Med Econ.* 2015;18(12):991-992

[6] Grim P. Rosenberger R, Eason RE et al. How simulations fail. *Synthese*. 2013;190(12):236-249

[7] Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *Inov Pharm*. 2016;7(2): No. 11.

[8] Langley PC. Nullius in verba: Version 2.0 of the University of Minnesota, School of Social and Administrative Pharmacy Program, Proposed Guidelines for Formulary Evaluation. *Inov Pharm.*2016;7(4): No.16

[9] Caro JJ, Briggs A, Siebert U et al. Modeling good research practices – Overview: A report of the ISPOR-SMDM Modeling good practices task force – 1. *Value Health*. 2012 (15:796-803

[10] Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices Task Force – 7. *Value Health*. 2012;15:843-850

[11] Langley PC**.** Imaginary worlds: Modeled claims for cost-effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *Inov Pharm*. 2016;7(2): No. 9

[12] Langley PC, Rhee TG. Imaginary worlds: A systematic review of the status of modeled cost-effectiveness claims published in the Journal of Medical Economics from January 2015 to December 2015. *Inov Pharm*. 2016;7(2): No. 16

[13] Langley PC, Rhee TG. Imaginary worlds: The status of modeled economic evaluation claims published in *Value in Health* January 2015 to November 2015. *Inov Pharm*. 2016;7(2): No. 18

[14] Langley PC, Rhee TG. Imaginary Worlds: The status of simulation modeling in claims for cost-effectiveness in diabetes mellitus. *Inov Pharm*. 2016;7(2): No. 17

[15] Langley PC, Rhee T. Imaginary Worlds: The Status of Modeled Quality Adjusted Life Year Claims for New Oral Anticoagulants in Atrial Fibrillation Published Between January 2012 and February 2016. *Inov Pharm.* 2016;7(3): No. 7

[16] National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: Nice, April 2013

[17] Zorginstituut Nederland. Guideline for economic evaluation in healthcare. Diemen: 2016

[18] Versteegh M, JKnies S, Brouwer W. From good to better: New Dutch guidelines for economic evaluations in healthcare. *Pharmacoeconomics*. 2016;34:1071-10743

[19] Langley PC**.** Sunlit uplands: the genius of the NICE reference case *Inov Pharm*. 2016;7(2): No.12

[20] Langley PC. Cost-Effectiveness and Formulary Evaluation: Imaginary Worlds and Entresto Claims in Heart Failure. *Inov Pharm.* 2016;7(3): No. 6

[21] Langley PC**.** Great expectations: Cost-utility models as decision criteria. *Inov Pharm*. 2016;7(2): No.14

[22] Hartung D, Bourdette D, Ahmed S et al. The cost of multiple sclerosis drugs in the US and the pharmaceutical industry. Too big to fail? *Neurology*. 2015:84:2185-2192